

**MHDO’s De-Identified Person Directory**

Last updated: 9/29/2022

# Purpose

The purpose of MHDO’s De-identified Person Directory is to provide a single set of person characteristics across MHDO ‘s data sets. The Directory (which is essentially a data table) will contain a set of attributes determined to be the most common valid value across all submitted records in MHDO’s APCD and MHDO’s Hospital Encounter Data. These attributes include date of birth, sex, race, and ethnicity. Authorized MHDO users will be able to link the data in the directory with data in the MHDO’s APCD and Hospital Encounter Data.

This document outlines the methodology used to determine the most common valid value for each of the attributes and the data elements that will be available in the Directory. The methodology will be reviewed annually and adjusted as needed.

# Methodology

The section below provides the goal and logic for creating the data elements that will populate the MHDO’s De-identified Person Directory for date of birth, sex, race, and ethnicity. The Directory will also contain the MHDO De-identified Person ID as a common identifier to allow linkage across MHDO Data Sets. Appendix A of this document contains a list data fields that will be available in the directory for release to approved data users, customized to their release requirements. This directory will be updated as new data is acquired through the standard submission process, with reassessment of methodology occurring on an annual basis.

## Date of Birth

The **goal** is to create data elements that contain the most common value for date of birth.

The **logic** for assignment is based on the most common valid value across all submitted records in MHDO’s APCD and MHDO’s Hospital Encounter Data.

## Sex

The **goal** is to create a composite structure that provides data users with information about the most common submitted sex, while also providing information on whether the sex changed across time.

The **logic** for creating these sex-related attributes and associated metadata elements is as follows:

* First Sex: Most common submitted sex across DW data sources (MHDO’s APCD, Hospital Encounter Data) in the *earliest* 6-month period for which this individual has data in the DW.
* Current Sex: For individuals with at least 12 months of data, most common submitted sex across DW data sources in the *most recent* available 6-month period. For individuals with < 12 months of data available, “N”.
* Sex\_Mixed\_Fl: Multiple sex values flag in the most recent 6 months of data. This flag is set to 1 if, in the *most recent* 6 months of data, at least 2 distinct values AND at least 30% of gender values <> Current\_Sex. This is an indication that this individual may be transgender.

## Race

The **goal** is to create a composite structure with as many fields as there are race and ethnicity categories in the OMB standard, rather than creating a single combined race and ethnicity field with mutually exclusive categories. This approach will provide more flexibility for data users.

The **logic** for creating a single composite set of race fields at the person ID level is to follow the hierarchy of data processing steps listed below, with each subsequent step filling in data for individuals that did not have data populated in one of the prior steps. This logic will be based on a rolling 5-year timeframe from the most recent data point available for each person ID to account for data collection changes or possible data errors.

1. The record sets identified as representing the de-prioritized scenarios listed below will be removed from the race and ethnicity observations used in the first steps for the composite. De-prioritized scenarios include:
2. Row level records where only ‘Other’, ‘Refused’, or ‘Unknown’ are available in the race/ethnicity data
   * 1. This decision was due to individuals having some rows of data with useful information (such as the race and ethnicity options that correspond to OMB Categories) along with other rows of data that only had unknown race/ethnicity, other race or refused to answer available. The goal of this is to prioritize the rows of data with useful race and ethnicity information and to supplement if no other information is available for them.
3. Records from Submitters that did not submit according to OMB categories, with least one record for each of the minority populations (Asian, American Indian/Alaska Native, Black, Native Hawaiian/Pacific Islander, Hispanic). This logic applies to both the APCD and Hospital Encounter datasets and is dependent on all of the data available within each data system.
   * 1. This decision was to account for the possibility that some submitters are not collecting all OMB categories as standalone categories. The goal of this is to prioritize submitters that have definitively reported each category over those that may not collect that information.
4. The remaining race and ethnicity observation records will be used to populate the fields in the Attribute Table data structure with the value 1 based on ever having one of the following racial or ethnic categories populated in either MHDO’S APCD or MHDO’s Hospital Encounter datasets (American Indian/Alaskan Native, Asian, Black/African American, Native Hawaiian/Pacific Islander, Hispanic, Non-Hispanic).
   * Note: A small proportion of records have conflicting ethnicity data (Hispanic and Non-Hispanic); in these instances, the logic will defer to Hispanic.
5. For individuals that did not have any race or ethnicity fields populated in the previous step, the removed observation records for scenario B will be used to populate the same race/ethnicity fields in the same manner as the previous step (i.e., each field will be populated based on ever having data for that racial or ethnic category).
   * Note: A small proportion of records have conflicting ethnicity data (Hispanic and Non-Hispanic); in these instances, the logic will defer to Hispanic.
6. For remaining individuals that had only ‘Other Race’ populated, Race\_Other will be populated with value 1.
7. For remaining individuals without any racial categories populated, individuals will have ‘Race Unknown’ populated with value 1. For remaining individuals without any ethnicity data populated, ‘Ethnicity Unknown’ will be populated with value 1.
8. For individuals that had evidence of multiple races at the eligibility month level in the MHDO’S APCD Source\_Multiracial\_Fl will be populated with value 1.
9. For individuals that have multiple races present within the composite, Inferred\_Multiracial\_Fl will be populated with value 1.
10. For individuals where the most recent data point is ‘Race Refused’, the Race\_refused\_Fl will be populated with value 1.

## Ethnicity

**Goal and Logic:** Listed above under Race.

# Appendix A: Data Fields

## Date of Birth

|  |  |
| --- | --- |
| Field Name | Description |
| DOB | Date of Birth (date) |
| DOB\_Day | Day element from date of birth (integer) |
| DOB\_Month | Month element from date of birth (integer) |
| DOB\_Year | Year element from date of birth (integer) |
| Multiple\_DOB\_fl | Multiple DOB values flag (numeric): 1 = yes; 0 = no. |

## Sex

| Field Name | Description |
| --- | --- |
| First\_Sex | Most common submitted sex across DW data sources in the *earliest* available 6-month period (char): M = male, F = female; U = unknown. |
| Current\_Sex | Most common submitted sex across DW data sources in the *most recent* available 6-month period (char): M = male, F = female; U = unknown; N = insufficient data. |
| Sex\_Mixed\_Fl | Flag for multiple sex values flag in the most recent available 6-month of data (numeric): 1 = yes; 0 = no. |

## Race

| Field Name | Description |
| --- | --- |
| AI\_AN | American Indian/Alaskan Native race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Asian | Asian race, composite assignment (numeric): 1 = yes, 0 = no. |
| Black | Black or African American race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| NH\_PI | Native Hawaiian/Pacific Islander race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| White | White or Caucasian race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Race\_Other | Other Race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Race\_Refused | Person Refused to answer Race, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Race\_Unk | Race is not otherwise known (numeric): 1 = yes, 0 = no. |
| Multiracial\_Fl | At least one DW observation indicates that the person identifies with multiple races (excluding when race is refused or unknown) (numeric): 1 = yes, 0 = no. |
| R\_Step\_Fl | Processing step Race was Derived, DW derived composite assignment (numeric): 1=1st Step, 2=2nd Step, 3=3rd Step, 4=4th Step |
| AI\_AN\_APCD\_EarliestDate | Earliest date the DW derived AI\_AN value appears in MHDO’S APCD (date) |
| AI\_AN\_ APCD\_LatestDate | Most recent date the DW derived AI\_AN value appears in MHDO’S APCD (date) |
| AI\_AN\_Hospital\_EarliestDate | Earliest date the DW derived AI\_AN value appears in MHDO’s Hospital Encounter data (date) |
| AI\_AN\_Hospital\_LatestDate | Most recent date the DW derived AI\_AN value appears in MHDO’s Hospital Encounter data (date) |
| Asian\_APCD\_EarliestDate | Earliest date the DW derived Asian value appears in MHDO’S APCD (date) |
| Asian\_APCD\_LatestDate | Most recent date the DW derived Asian value appears in MHDO’S APCD (date) |
| Asian\_Hospital\_EarliestDate | Earliest date the DW derived Asian value appears in MHDO’s Hospital Encounter data (date) |
| Asian\_Hospital\_LatestDate | Most recent date the DW derived Asian value appears in MHDO’s Hospital Encounter data (date) |
| Black\_APCD\_EarliestDate | Earliest date the DW derived Black value appears in MHDO’S APCD (date) |
| Black\_APCD\_LatestDate | Most recent date the DW derived Black value appears in MHDO’S APCD (date) |
| Black\_Hospital\_EarliestDate | Earliest date the DW derived Black value appears in MHDO’s Hospital Encounter data (date) |
| Black\_Hospital\_LatestDate | Most recent date the DW derived Black value appears in MHDO’s Hospital Encounter data (date) |
| NH\_PI\_APCD\_EarliestDate | Earliest date the DW derived NH\_PI value appears in MHDO’S APCD (date) |
| NH\_PI\_APCD\_LatestDate | Most recent date the DW derived NH\_PI value appears in MHDO’S APCD (date) |
| NH\_PI\_Hospital\_EarliestDate | Earliest date the DW derived NH\_PI value appears in MHDO’s Hospital Encounter data (date) |
| NH\_PI\_Hospital\_LatestDate | Most recent date the DW derived NH\_PI value appears in MHDO’s Hospital Encounter data (date) |
| White\_APCD\_EarliestDate | Earliest date the DW derived White value appears in MHDO’S APCD (date) |
| White\_APCD\_LatestDate | Most recent date the DW derived White value appears in MHDO’S APCD (date) |
| White\_Hospital\_EarliestDate | Earliest date the DW derived White value appears in MHDO’s Hospital Encounter data (date) |
| White\_Hospital\_LatestDate | Most recent date the DW derived White value appears in MHDO’s Hospital Encounter data (date) |
| Race\_Other\_APCD\_EarliestDate | Earliest date the DW derived Race\_Other value appears in MHDO’S APCD (date) |
| Race\_Other\_APCD\_LatestDate | Most recent date the DW derived Race\_Other value appears in MHDO’S APCD (date) |
| Race\_Other\_Hospital\_EarliestDate | Earliest date the DW derived Race\_Other value appears in MHDO’s Hospital Encounter data (date) |
| Race\_Other\_Hospital\_LatestDate | Most recent date the DW derived Race\_Other value appears in MHDO’s Hospital Encounter data (date) |
| Source\_Multiracial\_Fl\_APCD\_EarliestDate | Earliest date the DW derived Multiracial\_Fl value appears in MHDO’S APCD (date) |
| Source\_Multiracial\_Fl\_APCD\_LatestDate | Most recent date the DW derived Multiracial\_Fl value appears in MHDO’S APCD (date) |

## Ethnicity

| Field Name | Description |
| --- | --- |
| Hisp | Hispanic ethnicity, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Non\_Hisp | Non-Hispanic ethnicity, DW derived composite assignment (numeric): 1 = yes, 0 = no. |
| Ethn\_Unk | Ethnicity Unknown (numeric): 1 = yes, 0 = no. |
| E\_Step\_Fl | Step Race was Derived, DW derived composite assignment (numeric): 1=1st Step, 2=2nd Step, 3=3rd Step |
| Hisp\_APCD\_EarliestDate | Earliest date the DW derived Hisp value appears in MHDO’S APCD (date) |
| Hisp\_ APCD\_LatestDate | Most recent date the DW derived Hisp value appears in MHDO’S APCD (date) |
| Hisp\_Hospital\_EarliestDate | Earliest date the DW derived Hisp value appears in MHDO’s Hospital Encounter data (date) |
| Hisp\_ Hospital\_LatestDate | Most recent date the DW derived Hisp value appears in MHDO’s Hospital Encounter data (date) |
| Non\_Hisp\_APCD\_EarliestDate | Earliest date the DW derived Non\_Hisp value appears in MHDO’S APCD (date) |
| Non\_Hisp\_APCD\_LatestDate | Most recent date the DW derived Non\_Hisp value appears in MHDO’S APCD (date) |
| Non\_Hisp\_Hospital\_EarliestDate | Earliest date the DW derived Non\_Hisp value appears in MHDO’s Hospital Encounter data (date) |
| Non\_Hisp\_Hospital\_LatestDate | Most recent date the DW derived Non\_Hisp value appears in MHDO’s Hospital Encounter data (date) |